EP3260: Fundamentals of Machine Learning Over Networks

Homework 1

Group 3

Members: Jeannie He, Yusen Wang, Li Cheng, Yifei Dong

HW 1.1

A differential function is μ-strongly convex iff $\forall x_1, x_2 \in X$, $\mu > 0$.

$$f(x_2) \geq f(x_1) + \nabla f(x_1)^T (x_2 - x_1) + \frac{\mu}{2} \|x_2 - x_1\|_2^2 \quad *$$

Prove:

(1) $*$ is equivalent to a minimum positive curvature $\nabla^2 f(x) \geq \mu I_d$ $\forall x \in X$.

Assume that $f(x)$ is twice continuously differentiable.

Then its Taylor expansion can be written as:

$$f(x_2) = f(x_1) + \nabla f(x_1)^T (x_2 - x_1) + \frac{1}{2} (x_2 - x_1)^T \nabla^2 f(y) (x_1 - x_1)$$

① Necessity: $* \longrightarrow \nabla^2 f(x) \geq \mu I_d$

Proof: $(x_2 - x_1)^T \nabla^2 f(x) (x_2 - x_1) \geq \mu \|x_2 - x_1\|_2^2$ for any $x_1, x_2 \in \underline{X}$

Then, $(x_2 - x_1)^T (\nabla^2 f(y) - \mu I_d) (x_2 - x_1) \geq 0 \quad \forall x_1 \in \underline{X}, x_2 \in \underline{X}$

Since x_1 and x_2 are arbitrary,

Then $\nabla^2 f(x) \geq \mu I_d \quad \forall x \in \underline{X}$ must hold. Q.E.D.

② Sufficiency: $\nabla^2 f(x) \geq \mu I_d \longrightarrow *$

Proof: If $\nabla^2 f(x) \geq \mu I_d \quad \forall x \in \underline{X}$

Then, $f(x_2) = f(x_1) + \nabla f(x_1)^T (x_2 - x_1) + \frac{1}{2} (x_2 - x_1)^T \nabla^2 f(y) (x_2 - x_1)$

$$\geq f(x_1) + \nabla f(x_1)^T (x_2 - x_1) + \frac{\mu}{2} \|x_2 - x_1\|_2^2 \quad \forall x_1 x_2 \in \underline{X}$$

Thus, $f(x)$ is strongly convex. Q.E.D.

(2) $*$ is equivalent to $(\nabla f(x_2) - \nabla f(x_1))^T (x_2 - x_1) \geq \mu \|x_2 - x_1\|_2^2 \quad (*2)$

① Necessity: $* \longrightarrow *2$

If $f(x)$ is strongly convex, then

$$f(x_2) \geq f(x_1) + \nabla f(x_1)^T (x_2 - x_1) + \frac{\mu}{2} \|x_2 - x_1\|_2^2$$

$$f(x_1) \geq f(x_2) + \nabla f(x_2)^T (x_1 - x_2) + \frac{\mu}{2} \|x_2 - x_1\|_2^2$$

$$\Rightarrow 0 \geq (\nabla f(x_1) - \nabla f(x_2))^T (x_2 - x_1) + \mu \|x_2 - x_1\|_2^2$$

$$\Rightarrow (\nabla f(x_2) - \nabla f(x_1))^T (x_2 - x_1) \geq \mu \|x_2 - x_1\|_2^2$$

Thus, we have $* \longrightarrow *2$, Q.E.D.

② Sufficiency: $x_2 \to x^*$.

Proof $(\nabla f(x_2) - \nabla f(x_1))^T$

$(\nabla f(x_2) - \nabla f(x_1))^T (x_2 - x_1) \geq M (x_2 - x_1)^T (x_2 - x_1)$

$\Rightarrow ((\nabla f(x_2) - M x_2) - (\nabla f(x_1) - M x_1))^T (x_2 - x_1) \geq 0$

$g(x)$ is a convex function iff $(\nabla g(x_2) - \nabla g(x_1))^T (x_2 - x_1) \geq 0$

Let $g(x) = f(x) - \frac{M}{2}\|x\|^2$, Then

$(\nabla g(x_2) - \nabla g(x_1))^T (x_2 - x_1) \geq 0, \ \forall x_1, x_2 \in \bar{X}$

$\Rightarrow g(x)$ is convex $\Rightarrow g(x_2) \geq g(x_1) + \nabla g(x_1)^T (x_2 - x_1)$

$f(x_2) - \frac{M}{2}\|x_2\|^2 \geq f(x_1) - \frac{M}{2}\|x_1\|^2 + (\nabla f(x_1) - M x_1)^T (x_2 - x_1)$

$\Rightarrow f(x_2) \geq f(x_1) + \nabla f(x_1)^T (x_2 - x_1) + \frac{M}{2}(\|x_1\|^2 + \|x_2\|^2 - 2 x_1^T x_2)$.

$= f(x_1) + \nabla f(x_2)^T (x_2 - x_1) + \frac{M}{2}\|x_1 - x_2\|^2$

Thus, $x_2 \to x^*$, Q.E.D.

(3) Prove that * implies $f(x) - f^* \leq \frac{1}{2M}\|\nabla f(x)\|_2^2, \ \forall x$.

Proof: Take minization of x_2 on both sides of *.

Then, $f(x^*) \geq \min_{x_2 \in \bar{X}} \{ f(x_1)^T (x_2 - x_1) + \frac{M}{2}\|x_2 - x_1\|^2 \}$.

$= f(x_1) - \frac{1}{2M}\|\nabla f(x_1)\|_2^2$

Since x_1 is arbitrary, we have $f(x) - f^* \leq \frac{1}{2M}\|\nabla f(x)\|_2^2 \ \forall x \in \bar{X}$.

(4) Prove that * implies $\|x_2 - x_1\|_2 \leq \frac{1}{M}\|\nabla f(x_2) - \nabla f(x_1)\|_2, \ \forall x_1, x_2$.

Proof: In (2) we proof that * $\Leftrightarrow (\nabla f(x_2) - \nabla f(x_1))^T (x_2 - x_1) \geq M\|x_2 - x_1\|_2^2$

By Cauchy-Schwartz inequality, we have that

$\|\nabla f(x_2) - \nabla f(x_1)\| \cdot \|x_2 - x_1\| \geq \|(\nabla f(x_2) - \nabla f(x_1))^T \cdot (x_2 - x_1)\| \geq M\|x_2 - x_1\|_2^2$

$\Rightarrow \|\nabla f(x_2) - \nabla f(x_1)\| \geq M\|x_2 - x_1\| \quad \forall x_1, x_2 \in \bar{X}$

(5) Prove that * implies $(\nabla f(x_2)-\nabla f(x_1))^T(x_2-x_1) \le \frac{1}{M}\|\nabla f(x_2)-\nabla f(x_1)\|_2^2 \quad \forall x_1, x_2 \in \overline{X}$

Proof:

Let $g(y) = f(y) - \nabla f(x_1)^T \cdot y$.

Then we have:

$$(\nabla g(y_2) - \nabla g(y_1))^T(y_2-y_1) = (\nabla f(y_2)-\nabla f(y_1))^T(y_2-y_1) \ge M\|y_2-y_1\|^2$$

$\Rightarrow g(y)$ is strongly convex with respect to y.

In (3) we prove that: $g(x_2) - g^* \le \frac{1}{2M}\|\nabla g(x_2)\|^2 \quad \forall x_2$

Since $g^* = \min_{y \in \overline{X}} g(y) = g(x_1) = f(x_1) - \nabla f(x_1)^T \cdot x_1$.

Then $f(x_2) - f(x_1) - \nabla f(x_1)^T(x_2-x_1) \le \frac{1}{2M}\|\nabla f(x_2)-\nabla f(x_1)\|^2$.

$\Rightarrow f(x_1) - f(x_2) - \nabla f(x_2)^T(x_1-x_2) \le \frac{1}{2M}\|\nabla f(x_1)-\nabla f(x_2)\|^2$

$\Rightarrow (\nabla f(x_2)-\nabla f(x_1))^T(x_2-x_1) \le \frac{1}{M}\|\nabla f(x_2)-\nabla f(x_1)\|^2 \quad \forall x_1, x_2 \in \overline{X}$

(6) Prove that * implies $f(x)+r(x)$ is strongly convex for any convex f and strongly convex r.

Proof: f is convex $\Leftrightarrow f(x_2) \ge f(x_1) + \nabla f(x_1)^T(x_2-x_1) \quad \forall x_1, x_2 \in \overline{X}$

r is strongly convex $\Leftrightarrow r(x_2) \ge r(x_1) + \nabla r(x_1)^T(x_2-x_1) + \frac{M}{2}\|x_2-x_1\|^2 \quad \forall x_1, x_2$

Taking the sum of above equations:

$$r(x_2) + f(x_2) \ge f(x_1) + r(x_1) + (\nabla f(x_1) - \nabla r(x_1))^T(x_2-x_1) + \frac{M}{2}\|x_2-x_1\|^2$$

$$\Rightarrow h(x_2) \ge h(x_1) + \nabla h(x_1)^T(x_2-x_1) + \frac{M}{2}\|x_2-x_1\|^2$$

$\Rightarrow h(x) = f(x) + r(x)$ is strongly convex.

Problem 1.2 a)

We have: $\|\nabla f(x_2)+\nabla f(x_1)\|_2 \le L\|x_2-x_1\|_2$, $\forall x_1, x_2 \in \mathbb{R}^d$ (2)

We want to prove that (2) implies $f(x_2) \le f(x_1)+\nabla f(x_1)^T(x_1-x_2)+\frac{L}{2}\|x_2-x_1\|_2^2$

For this, let $g(\alpha)=f(x_1+\alpha(x_2-x_1))$, $\forall x_1, x_2 \in \mathbb{R}^d$; $\alpha \in [0,1]$. (i)

Then: $g'(\alpha)=\nabla f(x_1+\alpha(x_2-x_1))^T(x_2-x_1)$ and $g'(0)=\nabla f(x_1)^T(x_2-x_1)$. (ii)

$$g'(\alpha)=g'(0)+\nabla f(x_1+\alpha(x_2-x_1))^T(x_2-x_1)-g'(0)$$

$$=g'(0)+\nabla f(x_1+\alpha(x_2-x_1))^T(x_2-x_1)-\nabla f(x_1)^T(x_2-x_1)$$

$$=g'(0)+\left(\nabla f(x_1+\alpha(x_2-x_1))-\nabla f(x_1)\right)^T(x_2-x_1)$$

$$\le g'(0)+\|\nabla f(x_1+\alpha(x_2-x_1))-\nabla f(x_1)\|_2\|x_2-x_1\|_2$$

From (2), this gives: $g'(\alpha) \le g'(0)+L\|x_1+\alpha(x_2-x_1)-x_1\|_2\|x_2-x_1\|_2 =$

$$=g'(0)+L\|\alpha(x_2-x_1)\|_2\|x_2-x_1\|_2=$$

$$=g'(0)+\alpha L\|x_2-x_1\|\|x_2-x_1\|_2=g'(0)+\alpha L\|x_2-x_1\|_2^2$$

Hence, (2) implies $g'(\alpha) \le g'(0)+\alpha L\|x_2-x_1\|_2^2$ for $\alpha \in [0,1]$.

This gives $\int_0^1 g'(\alpha)\,d\alpha \le \int_0^1 g'(0)+\alpha L\|x_2-x_1\|_2^2\,d\alpha$

$$=g'(0)+L\|x_2-x_1\|_2^2\int_0^1 \alpha\,d\alpha=g'(0)+L\|x_2-x_1\|_2^2\left[\frac{\alpha^2}{2}\right]_0^1=$$

$$=g'(0)+\frac{L}{2}\|x_2-x_1\|_2^2$$

Hence, $g(1)=g(0)+\int_0^1 g'(\alpha)\,d\alpha \le g(0)+g'(0)+\frac{L}{2}\|x_2-x_1\|_2^2$

Since $g(1)=f(x_1+(x_2-x_1))=f(x_2)$; $g(0)=f(x_1)$; $g'(0)=\nabla f(x_1)^T(x_2-x_1)$, this gives:

$$\underbrace{f(x_2)}_{g(1)} \le \underbrace{f(x_1)}_{g(0)}+\underbrace{\nabla f(x_1)^T(x_2-x_1)}_{g'(0)}+\frac{L}{2}\|x_2-x_1\|_2^2$$ Q.E.D.

Problem 1.2b

We want to prove that (2) implies $f(x_2) \geq f(x_1) + \nabla f(x_1)^T (x_2 - x_1) + \frac{1}{2L} \|\nabla f(x_2) - \nabla f(x_1)\|_2^2$

b) Let $g(x) = f(x) - \nabla f(x_1)^T x$, $\forall x \in \mathbb{R}^d$. Then $\nabla g(x) = \nabla f(x) - \nabla f(x_1)$, so:

$$\|\nabla g(x_2) - \nabla g(x_1)\|_2 = \|(\nabla f(x_2) - \nabla f(x_1)) - (\nabla f(x_1) - \nabla f(x_1))\|_2 = \|\nabla f(x_2) - \nabla f(x_1)\|_2$$

By (2), we have $\|\nabla f(x_2) - \nabla f(x_1)\|_2 \leq L \|x_2 - x_1\|_2$ so this gives:

$$\|\nabla g(x_2) - \nabla g(x_1)\|_2 = \|\nabla f(x_2) - \nabla f(x_1)\|_2 \leq L \|x_2 - x_1\|_2.$$

From 1.2a), we have that if $\|\nabla g(x_2) - \nabla g(x_1)\|_2 \leq L \|x_2 - x_1\|_2$, then:

$$g(x_1) \leq g(x_2) + \nabla g(x_2)(x_1 - x_2) + \frac{1}{2} \|x_1 - x_2\|_2^2$$

Let now $tu = x_1 - x_2$ where $t \in \mathbb{R}$, $u \in \mathbb{R}^d$ and $\|u\| = 1$, then we have:

$$g(x_1) \leq g(x_2) + \nabla g(x_2)^T tu + \frac{L}{2} \|tu\|_2^2 = g(x_2) + t \nabla g(x_2)^T u + \frac{L}{2} t^2, \quad \forall x_1, x_2 \in \mathbb{R}^d$$

Since the equation above holds for any $x_1, x_2 \in \mathbb{R}^d$, we can have $t = -\frac{1}{L} \nabla g(x_2)^T u$ (note that this gives $g(x_1) = 0$), and get:

$$g(x_1) \leq g(x_2) - \frac{1}{L} (\nabla g(x_2)^T u)^2 + \frac{L}{2} \left(\frac{-1}{L} \nabla g(x_2)^T u\right)^2 =$$

$$= g(x_2) - \frac{1}{L} \|\nabla g(x_2)^T u\|_2^2 + \frac{1}{2L} \|\nabla g(x_2)^T u\|_2^2 = g(x_2) - \frac{1}{2L} \|\nabla g(x_2)^T u\|_2^2$$

$\leq$ {Cauchy-Schwartz inequality} $\leq g(x_2) - \frac{1}{2L} \|\nabla g(x_2)\|_2^2 \|u\|_2^2 =$

$= \{\|u\|_2^2 = 1\} = g(x_2) - \frac{1}{2L} \|\nabla g(x_2)\|_2^2$, so: $\boxed{g(x_1) \leq g(x_2) - \frac{1}{2L} \|\nabla g(x_2)\|_2^2}$

Since $g(x) = f(x) - \nabla f(x_1)^T x$ and $\nabla g(x) = \nabla f(x) - \nabla f(x_1)$, the above gives:

$$f(x_1) - \nabla f(x_1)^T x_1 \leq f(x_2) - \nabla f(x_1)^T x_2 - \frac{1}{2L} \|\nabla f(x_2) - \nabla f(x_1)\|_2^2, \text{ which gives:}$$

$$f(x_1) + \nabla f(x_1)^T (x_2 - x_1) + \frac{1}{2L} \|\nabla f(x_2) - \nabla f(x_1)\|_2^2 \leq f(x_2)$$

Hence: $\boxed{f(x_2) \geq f(x_1) + \nabla f(x_1)^T (x_2 - x_1) + \frac{1}{2L} \|\nabla f(x_2) - \nabla f(x_1)\|_2^2}$

Q.E.D

Problem 1.2c)

We want to prove that (2) implies $(\nabla f(x_2)-\nabla f(x_1))^T(x_2-x_1) \geq \frac{1}{L}\|\nabla f(x_2)-\nabla f(x_1)\|_2^2, \forall x_1, x_2 \in \mathbb{R}^d$

We have from 1.2b) that (2) implies $f(x_2) \geq f(x_1)+\nabla f(x_1)^T(x_2-x_1)+\frac{1}{2L}\|\nabla f(x_2)-\nabla f(x_1)\|_2^2, \forall x_1, x_2 \in \mathbb{R}^d$

Hence, (2) implies $-\nabla f(x_1)^T(x_2-x_1) \geq f(x_1)-f(x_2)+\frac{1}{2L}\|\nabla f(x_2)-\nabla f(x_1)\|_2^2$ (i)

By swapping x_1 & x_2, we also get: $-\nabla f(x_2)^T(x_1-x_2) \geq f(x_2)-f(x_1)+\frac{1}{2L}\|\nabla f(x_1)-\nabla f(x_2)\|_2^2$

This gives: $\nabla f(x_2)^T(x_2-x_1) \geq f(x_2)-f(x_1)+\frac{1}{2L}\|\nabla f(x_2)-\nabla f(x_1)\|_2^2$ (ii)

(i)+(ii) gives:

$-\nabla f(x_1)^T(x_2-x_1)+\nabla f(x_2)^T(x_2-x_1) \geq f(x_1)-f(x_2)+\frac{1}{2L}\|\nabla f(x_2)-\nabla f(x_1)\|_2^2+f(x_2)-f(x_1)+\frac{1}{2L}\|\nabla f(x_2)-\nabla f(x_1)\|_2^2 \Leftrightarrow$

$\Leftrightarrow -\nabla f(x_1)^T(x_2-x_1)+\nabla f(x_2)^T(x_2-x_1) \geq \frac{1}{L}\|\nabla f(x_2)-\nabla f(x_1)\|_2^2 \Leftrightarrow$

$\Leftrightarrow (\nabla f(x_2)-\nabla f(x_1))^T(x_2-x_1) \geq \frac{1}{L}\|\nabla f(x_2)-\nabla f(x_1)\|_2^2$

Q.E.D.

Problem 1.3 Suppose there is a sequence $\{x_k\}$ that converges to some point x, and $\lim_{k\to\infty} \frac{|x_{k+1}-x|}{|x_k-x|^r} = C < \infty$ for some $r, C > 0$ (1.3)

a) • If $r = C = 1$, then the convergence rate of the sequence is sublinear.

• While being slower than linear, superlinear and quadratic convergence rate, sublinear convergence rate has the benefit of allowing algorithms to converge slow enough to ensure high precision on problems where fast convergence may lead to the algorithm missing the optimal solution, such as when the data has too high volume and/or complexity.

• Example: $x_k = \frac{1}{k}$

Proof: We have $x = \lim_{k\to\infty} x_k = \lim_{k\to\infty} \frac{1}{k} = 0$ and

$$\lim_{k\to\infty} \frac{|x_{k+1}-x|}{|x_k-x|^r} = \lim_{k\to\infty} \frac{\left|\frac{1}{k+1}-0\right|}{\left|\frac{1}{k}-0\right|^r} = \lim_{k\to\infty} \frac{\frac{1}{k+1}}{\left(\frac{1}{k}\right)^r} = \lim_{k\to\infty} \frac{k^r}{k+1} = C$$

From the above, we see that $C < \infty$ only if $r \le 1$ and that $r = 1 \Rightarrow C = 1$

b) • If $r = 1$ and $C < 1$, then the convergence rate is called linear.

• Linear convergence has the benefit of being faster than sublinear convergence rate. Since it is slower than superlinear and quadratic convergence rate, it is suitable when time is equally or slightly less important than precision or when fast convergence can easily lead to the algorithm converging towards "wrong" convergence point.

• Example: $x_k = 1 + \left(\frac{1}{2}\right)^k$

Proof: We have $x = \lim_{k\to\infty} x_k = \lim_{k\to\infty} 1 + \left(\frac{1}{2}\right)^k = 1$ and

$$\lim_{k\to\infty} \frac{|x_{k+1}-x|}{|x_k-x|^r} = \lim_{k\to\infty} \frac{\left|1+\left(\frac{1}{2}\right)^{k+1}-1\right|}{\left|1+\left(\frac{1}{2}\right)^k-1\right|^r} = \lim_{k\to\infty} \frac{\left(\frac{1}{2}\right)^{k+1}}{\left(\left(\frac{1}{2}\right)^k\right)^r} = \lim_{k\to\infty} \frac{\left(\frac{1}{2}\right)^{k+1}}{\left(\frac{1}{2}\right)^{kr}} = C$$

where $C < \infty$ only if $r \le 1$ and $r = 1 \Rightarrow C = \frac{1}{2} < 1$

c) & d) on next page!

Problem 1.3 c+d

c) • The convergence rate of $\{x_k\}$ is defined as being superlinear if we have $r=1$ and $c=0$ in (1.3).

• Superlinear convergence rate is faster than linear convergence rate but slower than quadratic convergence rate. It is therefore useful on problems where accuracy and time to convergence are both important, or when time to convergence is desired but that too fast convergence may lead to the algorithm "missing" the convergence point, or when it is difficult to achieve faster convergence rate.

• Example: $x_k = \frac{1}{k^k}$ as it gives: $\lim_{k\to\infty} x_k = \lim_{k\to\infty} \frac{1}{k^k} = 0$ and $x = \lim_{k\to\infty} x_k$ gives:

$$\lim_{k\to\infty} \frac{|x_{k+1}-x|}{|x_k-x|^r} = \lim_{k\to\infty} \frac{\left|\frac{1}{(k+1)^{k+1}}-0\right|}{\left|\frac{1}{k^k}-0\right|^r} = \lim_{k\to\infty} \frac{\frac{1}{(k+1)^{k+1}}}{\left(\frac{1}{k^k}\right)^r} = \lim_{k\to\infty} \frac{k^{kr}}{(k+1)^{k+1}} = C$$

where $C<\infty$ only if $r\leq 1$ where $r=1$ gives $C=0$.

d) • The convergence rate of $\{x_k\}$ is defined as quadratic if $r=2$ in (1.3).

• Quadratic convergence rate has the advantage of being faster than the aforementioned convergence rates. With quadratic convergence rates, the time and number of iterations required to reach a given threshold is expected to be less than with the aforementioned convergence rates. This makes it suitable for problems where time is important.

• Example: $x_k = e^{-2^k}$

Proof: $\lim_{k\to\infty} e^{-2^k} = 0$ so $x = \lim_{k\to\infty} x_k$, $x_k = e^{-2^k}$ gives:

$$\lim_{k\to\infty} \frac{|x_{k+1}-x|}{|x_k-x|^r} = \lim_{k\to\infty} \frac{|e^{-2^{k+1}}|}{|e^{-2^k}|^r} = \lim_{k\to\infty} \frac{e^{-2^{k+1}}}{(e^{-2^k})^r} = \lim_{k\to\infty} \frac{e^{-2^k\cdot 2}}{e^{-2^k r}} = \lim_{k\to\infty} \frac{e^{2^k r}}{e^{-2^k\cdot 2}} = C$$

where $C<\infty$ only if $r\leq 2$ and $r=2$ gives $C=1$.

Problem 1.4

February 5, 2023

1 Problems

1.1 a

Newton's method is a good fit for such a problem with a relatively small dimension. The Hessian matrix could be efficiently calculated numerically with a complexity of N^2. Compared with gradient descent, it will converge faster.

1.2 b

For such a problem with high dimensionality, the second-order derivative requires too much computation effort. As a better practice, stochastic gradient descent or mini-batch method avoids in a sense the whole batch problem and saves a lot of computation in each iteration.

1.3 c

As mentioned in problem b, Newton's method will not work due to efficiency issues. In the case of $p = 1, b = 1$, a proximal algorithm such as FISTA could be a good choice. We formulate the problem as minimizing $f(x) + g(x)$ where f is the original term to minimize and g is the convex indicator function of the probability simplex. For efficient calculation of the Hessian matrix H, we could consider approximating it by $J^T J$, which is readily proved by Taylor expansion around a certain point x_0 (in the current loop), $H = J^T J$ approximated holds for points around x_0.

1.4 d

Since $r(\mathbf{x})$ could be possibly non-convex, we need to consider non-convex optimization for the new problem. Alternating minimization and Non-Convex Projected Gradient Descent are two algorithms that can be applied here. $r(\mathbf{x})$, on the other hand, could be convex as well. The sum of $N + 1$ convex functions remains a convex, and thus we say the results above hold.

A Appendix

As a complementary material, we wrote a Google Colab script in which we run convex optimization over a naive quadratic programming problem using OSCP solver [SBG+20]. The solver runs the ADMM algorithm (for more details see the cited paper). We run over p from 1 to 10, and N from 1 to 10^4 to see how the parameters of dimensionality affect the solver's speed. The number of equality constraints p does not seem to affect the time complexity a lot, since they are all encompassed in the loss function in the augmented-Lagrangian approach. From the simple printouts, we can see how N affects the computational speed in a nonlinear way.

References

[SBG+20] Bartolomeo Stellato, Goran Banjac, Paul Goulart, Alberto Bemporad, and Stephen Boyd. Osqp: An operator splitting solver for quadratic programs. *Mathematical Programming Computation*, 12(4):637–672, 2020.

HW1.5: We have:

① $f(x)$ is L-Smooth

② $f(x)$ is μ-strongly convex

To prove:

$$[\nabla f(x)-\nabla f(y)]^T(x-y)\geq \frac{\mu L}{\mu+L}\|x-y\|_2^2+\frac{1}{\mu+L}\|\nabla f(x)-\nabla f(y)\|_2^2$$

Proof: Define $g(x)=f(x)-\frac{\mu}{2}\|x\|_2^2$. Then $\nabla g(x)=\nabla f(x)-\mu x$.

Since ①, we have: $\|\nabla f(x)-\nabla f(y)\|_2\leq L\|x-y\|_2$. Then:

$$\begin{aligned}\|\nabla g(x)-\nabla g(y)\|_2^2 &= \|\nabla f(x)-\nabla f(y)-\mu(x-y)\|_2^2\\ &= \|\nabla f(x)-\nabla f(y)\|_2\left(\|\nabla f(x)-\nabla f(y)\|_2-2\mu\|x-y\|_2\right)\\ &\quad +\mu^2\|x-y\|_2^2\\ &\leq L\|x-y\|_2\left(L\|x-y\|_2-2\mu\|x-y\|_2\right)+\mu^2\|x-y\|_2^2\\ &= (L-\mu)^2\|x-y\|_2^2\end{aligned}$$

So that $g(x)$ is $(L-\mu)$-smooth.

With the conclusion in HW 1.2, we have:

$$[\nabla g(x)-\nabla g(y)]^T(x-y)\geq\frac{1}{L-\mu}\|\nabla g(x)-\nabla g(y)\|_2^2$$

$$\Rightarrow [\nabla f(x)-\nabla f(y)-\mu(x-y)]^T(x-y)\geq\frac{1}{L-\mu}\|\nabla f(x)-\nabla f(y)-\mu(x-y)\|_2^2$$

After Expanding and rearranging items, we have:

$$\left(1+\frac{2\mu}{L-\mu}\right)[\nabla f(x)-\nabla f(y)]^T(x-y)\geq\left(\mu+\frac{\mu^2}{L-\mu}\right)\|x-y\|_2^2+\frac{1}{L-\mu}\|\nabla f(x)-\nabla f(y)\|_2^2$$

$$\Rightarrow \frac{L+\mu}{L-\mu}[\nabla f(x)-\nabla f(y)]^T(x-y)\geq\frac{\mu L}{L-\mu}\|x-y\|_2^2+\frac{1}{L-\mu}\|\nabla f(x)-\nabla f(y)\|_2^2$$

$$\Rightarrow [\nabla f(x)-\nabla f(y)]^T(x-y)\geq\frac{\mu L}{\mu+L}\|x-y\|_2^2+\frac{1}{\mu+L}\|\nabla f(x)-\nabla f(y)\|_2^2$$

Q.E.D.